1. Motivation
2. Previous Work
3. Challenges

Motivation

Importance of White Blood Cell Count

White blood cells (WBCs), also called leukocytes, are an integral part of the human immune system. These cells fight infections by attacking bacteria, viruses, and other foreign pathogens that invade the body. Having too many or too few white blood cells can indicate a blood disorder. Disorders affecting these cells often result in the body's inability to eliminate or control infections. [1] Therefore, WBC count is of interest in clinical pathology and oncology.

A general WBC count can detect hidden infections within the body and alert doctors to undiagnosed medical conditions, such as autoimmune diseases, immune deficiencies, blood disorders, and cancer. However, a WBC count can indicate that there is a disease or condition affecting white blood cells, but it cannot determine the underlying cause. Several other tests may be performed at the same time or in follow up to an abnormal result to help make a diagnosis. One of these tests is a blood smear review. [2]

Blood Smear Test

A blood smear is a blood test used to look for abnormalities in red blood cells, WBCs, and platelets. The test provides information on the number and shape of these cells, which can help doctors diagnose certain medical conditions. There are five distinct subtypes of WBCs each with a unique nucleus morphology and function.

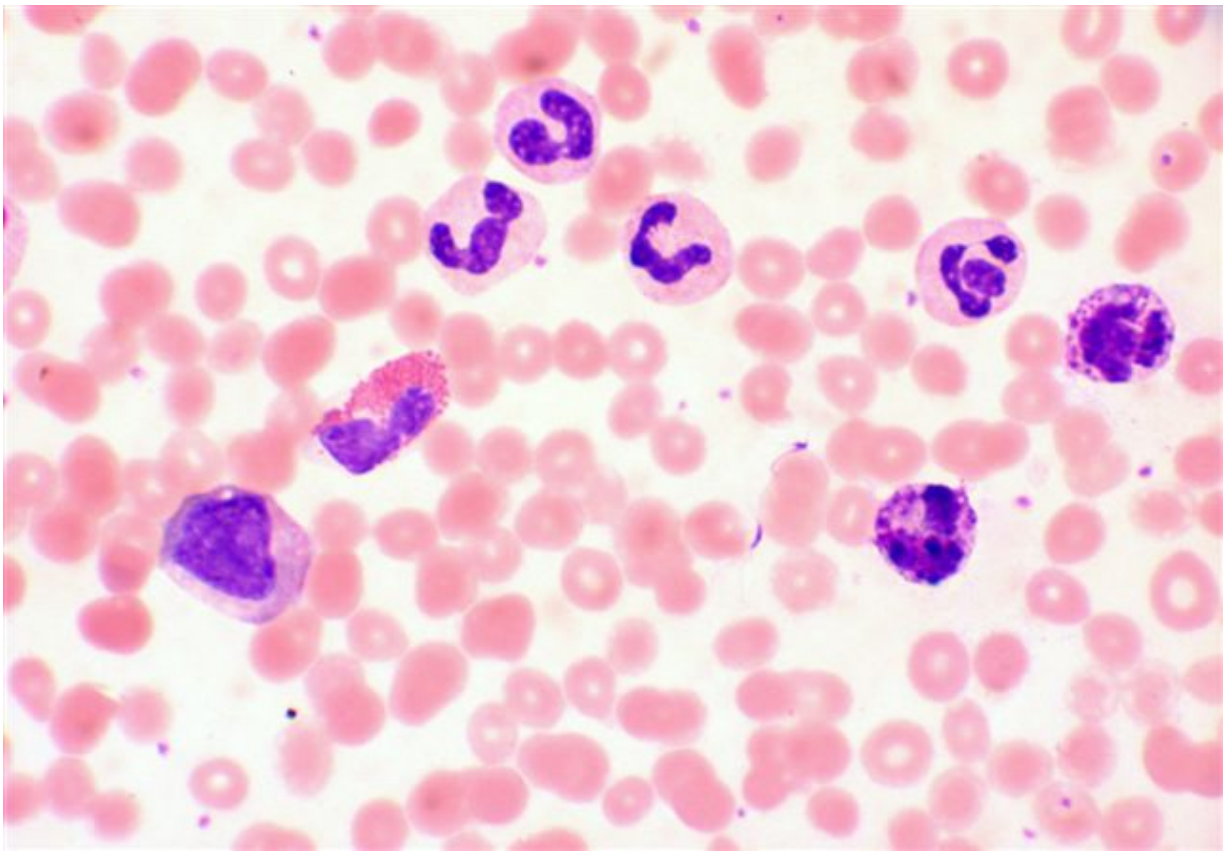


Figure 1. Example of stained blood smear slide. [3]

When stained, these morphological differences become more dramatic and can be seen in Table 1. Abnormalities in the shape or number of WBCs may be signs of the diseases mentioned above. Therefore, it is extremely important to obtain an accurate count of each subtype if any immune-linked disorders are suspected.

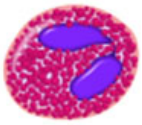
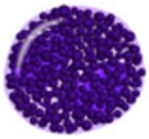
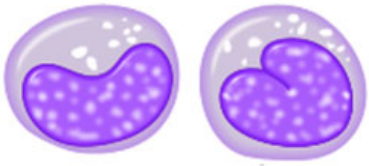
Subtype	Nucleus	Function	Example
Neutrophil	Multi-Lobed	Bacterial or fungal infection. These are the most common first responders to microbial infection.	
Eosinophil	Bi-Lobed	Parasitic infections and allergic reactions (inflammatory).	
Basophil	Bi/Tri-Lobed	Allergic and antigen response (releases histamine causing vasodilation).	
Lymphocyte	Deep Staining, Eccentric	Include B cells, CD4+ helper T cells, and CD8+ cytotoxic T cells. Operate primarily in the lymphatic system.	
Monocyte	Kidney Shaped	Phagocytosis of pathogens. Presentation of antigens to T cells. Eventually, they become tissue macrophages, which remove dead cell debris and attack microorganisms.	

Table 1. White blood cell subtypes, functions, and morphology. [1]

Accurate cell counting is a long and tedious process. Scientists and pathologists individually analyze each cell’s morphology through light microscopy. This can take up to several hours, time that a patient may not have. In addition, since it is a tiring process, observer fatigue can produce inaccurate results. We aim to quantify the morphological differences between the WBC subtypes and classify them in order to aid pathologists. An automatic isolation and classification algorithm would significantly decrease the time needed for a blood smear analysis, leading to faster, potentially life saving results.

Previous Work

Challenges

We are attempting to classify and count the different subspecies of WBC in multi-cell images using morphological features. This is a challenging and ongoing problem explored by many independent researchers, and there is no standard solution. While shape and feature recognition is easy to humans, it is one of the most researched topics in machine learning. Most methods divide the process into two subsystems, segmentation and classification.

Past Methods

Segmentation

We must segment each cell image into the nucleus and cell body in order to extract features. The first problem is cell isolation. Most researchers cropped the individual cells manually due to the difficulty in isolating overlapping circles (WBCs and red blood cells) which can be seen in Figure 1. Several methods for segmentation have been explored in the literature.

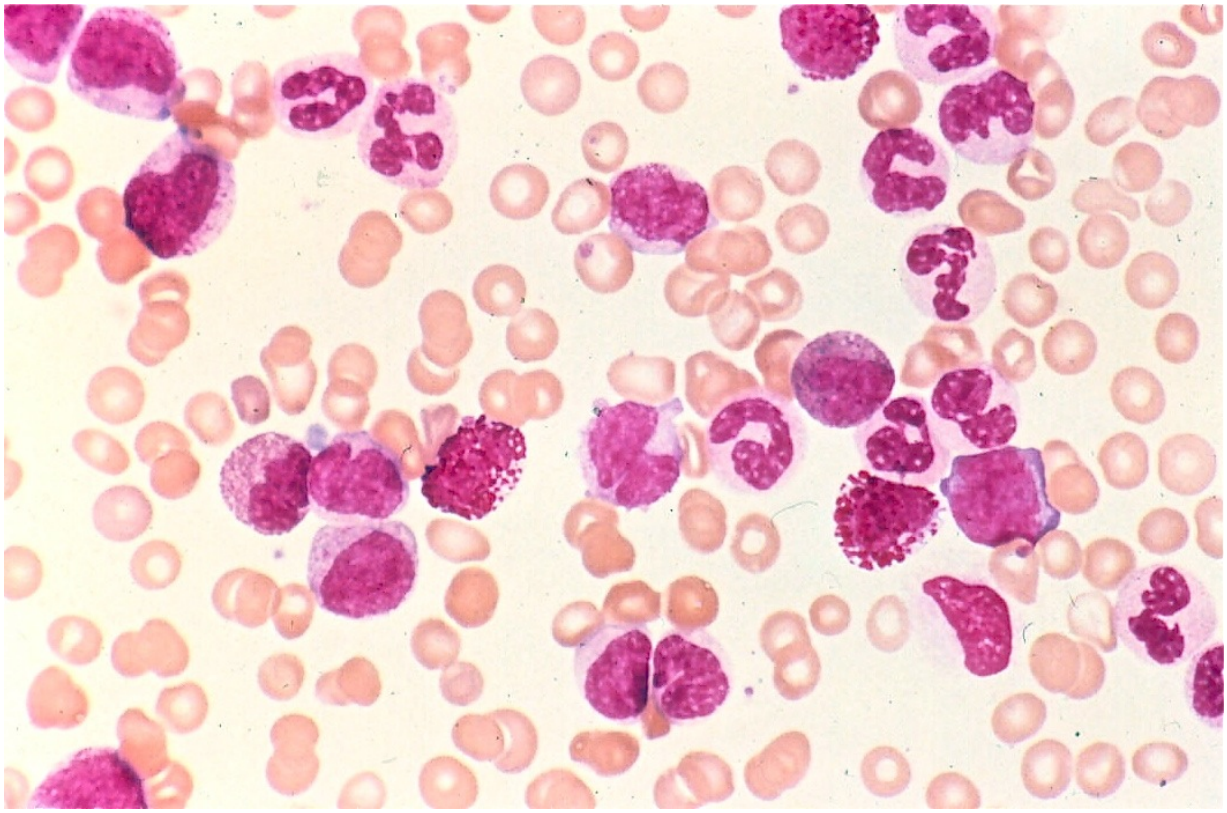


Figure 1. Overlapping can be seen between RBCs and WBCs.

Green channel filtering is a common way to isolate WBCs in colored blood smear images. Red blood cells (RBCs) and white blood cells separate into two distinct peaks in this space (with number of pixels on the y axis and color value, 0 to 255 on the x axis). The white blood cells have more green content. Furthermore, the WBC nuclei also differ slightly from the cytoplasm in green content. [1] However, this method requires extremely consistent and high quality images due to the subtleties associated with the different cells' green channel content.

Grayscale images and thresholding can also be used to separate the RBCs, WBCs, and WBC nuclei. WBC cell bodies appear much darker than RBCs, and WBC nuclei appear darker than cell bodies. Several thresholds can be used to differentiate between the cells. [2] For example, in "White Blood Cell Segmentation Using Morphological Operators and Scale-Space Analysis," a hard-coded threshold was used to isolate the first seed. Then, another equation was used to improve the gradient extracted. This was repeated for 10 iterations. [3] This method requires the images to be high

resolution, because blur and graininess can cause individual areas of darker or lighter spots.

Some researchers use a combination of both common methods in order to obtain the best results. They first localize the nuclei using color analysis, and then threshold for finer boundaries. [1]

Past Classification Methods

We found that many research articles used both morphological and texture features for classification. These are based mostly on nucleus shape. Recall that the subtypes of WBCs have lobed nuclei which differ between the types. In the literature, monocytes are the most difficult to classify, often mistaken lymphocytes and neutrophils.

Most research papers implement at least ten features. These features must differ significantly between the cell types. They include a variety of area, perimeter, and color characteristics. The papers that used colored images have more features than grayscale due to the differences in color content as mentioned above. Below are example categories that have been explored.

Intensity and color based features

- Mean and variance of color distances for cell, cytoplasm, and nucleus

Texture based features

- Contrast of cytoplasm and nucleus
- Homogenicity of cytoplasm and nucleus
- Entropy of cytoplasm and nucleus

Shape based features

- Area of cell and nucleus
- Ratio of nucleus area over cell area and perimeter length
- Nucleus shape features
- Number of nuclei

Common classification methods include k-nearest neighbors, learning vector quantization, Multi-layer perceptrons, and support vector machine (SVM). These are pattern recognition algorithms, which makes sense for this application. We found that the most successful paper, which used 57 features in total from all three categories above, had an accuracy of 91% for SVM. [3]

Challenges

`asdfasdfhakjsdhfksjdf`

We are attempting to classify and count the different subspecies of WBC in multi-cell images. Therefore, we must divide the process into two subsystems. First, we have to isolate each individual cell from each image. This will be done using centroid detection and masking because in stained cell images, the nucleus of the cell is the most prominent and high contrast feature. We make the assumption that isolating the nucleus leads to isolating the cell. We use a multi-threshold method to isolate the cell. The goal of the isolation algorithm is to accurately separate the cell body from the background and the cell nucleus from the cell body with the appropriate curved boundaries. (insert image of mask) Each isolated cell body and nucleus will be saved as a matrices in grayscale and binary. This results in a total of four processed images per original image. These are needed to extract features relevant to the classification of WBC subtypes. The features include morphological characteristics of both the cell and nuclei. We will implement both support vector machine (SVM) and neural network algorithms with these features. The algorithms are trained with prototypical images of the five cell types and used to classify the cells. Because WBC nuclei are morphologically distinct, we expect that this method will be able to differentiate between the different types. This is challenging because while shape and feature recognition is easy to humans, it is one of the most researched topics in machine learning. (insert example neural network) After each cell is classified, we will count the number of each subtype and compare against human results (which one of us will pain-stakingly do). We will test 20 multi-cell images in total to generate accuracy and precision statistics.

(1,1)	(1,2)	(1,3)
(2,1)	(2,2)	(2,3)

(3,1)	(3,2)	(3,3)